# Real-time PCR

**M. Tevfik Dorak (Ed.)**

School of Clinical Medical Sciences (Child Health)
Newcastle University
Newcastle-upon-Tyne, UK

Taylor & Francis
Taylor & Francis Group

Published by:

Taylor & Francis Group

In US: 270 Madison Avenue
New York, N Y 10016
In UK: 2 Park Square, Milton Park
Abingdon, OX14 4RN



ISBN: 0-4153-7734-X



A catalog record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data

Real-time PCR/M. Tevfik Dorak (ed.).
p. cm.
Includes bibliographical references and index.
ISBN 1-415-37734-X (alk. paper)
1. Polymerase chain reaction. I. Dorak, M. Tevfik.
[DNLM: 1. Polymerase Chain Reaction – methods. QU 450 R2874 2006]
QP 606.D46R43 2006
572'.43–dc22

Editor: Elizabeth Owen
Editorial Assistant: Kirsty Lyons
Production Editor: Karin Henderson
Typeset by: Phoenix Photosetting, Chatham, Kent, UK
Printed by: MPG BOOKS Limited, Bodmin, Cornwall, UK

Printed on acid-free paper

10 9 8 7 6 5 4 3 2 1

T&F informa
Taylor & Francis Group, an informa business

Visit our web site at http://www.garlandscience.com

# Conte

# 5 High-throughput primer and probe design

*Xiaowei Wang and Brian Seed*

The design of appropriate primers for the specific quantitative generation of DNA amplicons from diverse transcripts is an important requirement for most applications of real-time PCR. For the simultaneous measurement of multiple transcripts in a single experiment, an additional constraint, that all reactions proceed efficiently under the same conditions, must be imposed. In the first section of this chapter, we describe general primer and probe design guidelines for real-time PCR. In the second part, we present an online real-time PCR primer database encompassing most human and mouse genes identified to date. The primer database contains primers that perform well under a single set of conditions, allowing many simultaneous determinations of mRNA abundance to be carried out.

## 5.1 Primer and probe design guidelines

### 5.1.1 Primer specificity

Non-specific amplification is one of the greatest challenges for the successful deployment of real-time PCR methods intended to be used for the validation or discovery of transcript abundance variation. In addition to the sequence of interest, many thousands of other sequences can be expected to be present in such applications. The design of PCR primers for this purpose should therefore take into account the potential contribution of all possible off-target template sequences, in order to prevent mispriming. This is usually achieved by comparing sequence similarity between the primers and all other template sequences in the design space.

### 5.1.2 Primer length

PCR primers are typically 16–28 nucleotides long. If the length is too short, it is difficult to design gene-specific primers and choose optimal annealing temperature. On the other hand, very long oligos unnecessarily increase oligo synthesis cost. In addition, longer primers are more likely to form secondary structures that result in decreased PCR efficiency or promote primer dimer formation, since the primers constitute the nucleic acid sequences at highest concentration in the reaction.

### 5.1.3 Primer GC content

In most PCR applications the primer GC content lies between 35% and 65%. If the GC content is too high, mispriming frequently results, because

even a short stretch of oligo sequence may form a stably annealed duplex with non-target templates. On the other hand, very low GC content may result in poor primer binding, leading to decreased PCR efficiency.

### 5.1.4 Primer 3' end stability

The 3' end residues contribute strongly to non-specific primer extension by Taq DNA polymerase, especially if the binding of these residues is relatively tight to the non-target template. Therefore, primers with very high 3' terminal stability should be rejected. The binding stability can be calculated from the free energy profile (ΔG). Typically the more computationally demanding aspects of calculating free energy, such as loop entropy, can be ignored because of the unfavorable energetics of opening small loops (internal denaturation) compared to end melting.

### 5.1.5 Primer sequence complexity

Low-complexity sequences should be identified and discarded during primer design, due to the likelihood of mispriming with such primers. In addition, some types of low sequence complexity pose a challenge for oligonucleotide synthesis. For example, a large number of contiguous guanosine residues in a primer can lead to poor synthesis yield due to decreased chemical coupling efficiency. In addition the resulting primers can exhibit poor solubility in aqueous media.

### 5.1.6 Primer melting temperature

The melting temperature (Tm) is the most important factor in determining the optimal PCR annealing temperature. An ideal PCR reaction should have forward and reverse primers with similar Tm values. Tm is not only determined by primer sequence, but also by other parameters, such as salt concentration and primer concentration. In recent years, extensive thermodynamic studies have been carried out to accurately determine oligo Tm values.

Currently, the following methods for Tm calculation are adopted by most primer design programs.

#### The '4 + 2' rule

$$Tm = 4 * (G + C) + 2 * (A + T).$$

This is a simple equation solely based on primer GC content. Tm is calculated by counting the total number of G/C and A/T. Each G/C contributes 4°C and each A/T contributes 2°C to Tm. This method is sometimes used to quickly approximate Tm values for very short oligos (<10 residues). However PCR primers are usually much longer and this Tm calculation method is not generally recommended.

#### Simple equation based on GC content and salt concentration

$$Tm = 81.5 + 16.6 * \mathrm{Log}_{10}[Na^{+}_{eq}] + 0.41 * (\%GC) - 600/L$$

where $[Na^{+}_{eq}]$ is the eq
percent of G/C in the
1989). This method is
because it is relatively
calculated in this way
degrees different from
method is considered t
long DNA duplex, such

#### The nearest neighbc

where R is the gas con
tion, $\Delta H^{\circ}$ is the enthal
*et al.*, 1986). This is cor
oligo thermodynamic
primer Tm determinati
determined thermodyr
(SantaLucia, Jr., 1998).

The entropy change
an entropy correction
$\ln[Na^{+}_{eq}]$, where L is the
molar concentration fr
study (von Ahsen *et al.*
equation:

$$[Na^{+}_{eq}] = [Mo$$

where monovalent cati
buffer.

### 5.1.7 Primer location

Oligo dT and random p
(RT) to produce cDNA t
specifically to mRNA
products. However, this
sis because it is difficul
extension capability. If
primers should be pick
maximum assay sensitiv

In contrast, random p
age. Random primers
sequence. Because all
transcript, there will be
highest cDNA abundan
adopted, real-time PCR
target sequence for max

where $[Na^+_{eq}]$ is the equivalent sodium molar concentration, (%GC) is the percent of G/C in the primer, and L is the primer length (Sambrook *et al.*, 1989). This method is widely used in many computer programs, partially because it is relatively easy to implement. Despite its simplicity, Tm values calculated in this way usually have reasonable accuracy and are only a few degrees different from empirically determined Tm values. In addition, this method is considered to be the most accurate way to calculate Tm for very long DNA duplex, such as PCR amplicon.

### The nearest neighbor method

$$Tm = \Delta H^\circ / (\Delta S^\circ - R \ln (C_T/4))$$

where R is the gas constant (1.987 cal/Kmol), $C_T$ is the primer concentration, $\Delta H^\circ$ is the enthalpy change, and $\Delta S^\circ$ is the entropy change (Breslauer *et al.*, 1986). This is considered to be the most accurate method to calculate oligo thermodynamic stability, and thus is the recommended method for primer Tm determination. $\Delta H^\circ$ and $\Delta S^\circ$ are calculated using the empirically determined thermodynamic parameters for neighboring bases in a primer (SantaLucia, Jr., 1998).

The entropy change is significantly affected by salt concentration. Thus, an entropy correction is required: $\Delta S^\circ = \Delta S^\circ$ (1 M $Na^+$) + 0.368 (L – 1) $\ln[Na^+_{eq}]$, where L is the primer length and $[Na^+_{eq}]$ is the equivalent sodium molar concentration from all salts in a PCR reaction. According to a recent study (von Ahsen *et al.*, 2001), $[Na^+_{eq}]$ can be determined by the following equation:

$$[Na^+_{eq}] = [\text{Monovalent cations}] + 120 \sqrt{[Mg^{2+}] - [\text{dNTPs}]},$$

where monovalent cations are typically present as $Na^+$, $K^+$ and $Tris^+$ in PCR buffer.

## 5.1.7 Primer location in the sequence

Oligo dT and random primers are commonly used in reverse transcription (RT) to produce cDNA template for real-time PCR. Oligo dT primers anneal specifically to mRNA poly(A) tails, thus minimizing non-coding cDNA products. However, this priming strategy introduces 3′ bias in cDNA synthesis because it is difficult to produce full-length cDNAs due to limited RT extension capability. If oligo dT primers are used in RT, the real-time PCR primers should be picked from the 3′ region of a gene sequence to gain maximum assay sensitivity.

In contrast, random primers are often used in RT for full transcript coverage. Random primers can potentially bind to any site in a transcript sequence. Because all primers are extended toward the 5′ end of the transcript, there will be a linear gradient of sequence representation, with highest cDNA abundance in the 5′ regions. If random priming strategy is adopted, real-time PCR primers should be picked close to the 5′ end of the target sequence for maximum sensitivity in real-time PCR.

### 5.1.8 Amplicon size

PCR efficiency can be affected by amplicon size. Very long amplicons leads to decreased PCR efficiency. Since PCR efficiency is one of the most important factors for accurate expression quantification, the amplicon should be smaller than 250 bp. Typically the size range is 100–250 bp.

### 5.1.9 Cross-exon boundary

To minimize the effect of DNA contamination in RNA template, the forward and reverse primers can be designed from different exons and to span exon–intron boundaries. In this way one can reduce the genomic DNA contribution to expression quantitation. This design strategy is most important for accurate quantitation of low-expressing genes or genes with many loci in the genome. However, in general, DNA contamination has minimal effect on real-time PCR because a typical transcript copy number is much higher than the number of gene loci and standard RNA preparation procedures remove genomic DNA efficiently. In any case, this strategy may fail if pseudogenes of the target gene are present in the genome.

### 5.1.10 Primer and template sequence secondary structures

Primer or target template secondary structures can retard primer annealing, leading to decreased amplification efficiency. The likelihood of secondary structure is greatest in regions rich in complementary base pairing, such as the stem of a stem-loop structure (Mount, 2001). If part of a primer or target sequence is inaccessible due to secondary-structure formation, the primer annealing efficiency may decrease dramatically.

In addition, primer secondary-structure may lead to primer dimer formation, which is one of the biggest challenges for accurate quantification by real-time PCR, especially when DNA intercalating dyes (e.g., SYBR® Green I) are used. Primer dimers can be produced by primer self-annealing, or by annealing between the forward and reverse primers. Therefore, the forward and reverse primers should be evaluated together during design to avoid potential primer dimer formation.

### 5.1.11 TaqMan® probe design

Many of the guidelines for primer design are also applicable to TaqMan® probe design. It is recommended to use Primer Express® software (Applied Biosystems) for TaqMan® assay design.

- The probe melting temperature in general should be ~10°C higher than the forward or reverse primer.
- Do not put G at the 5′ end of the probe as this will quench reporter fluorescence.
- In general the GC content should be 35–65%.
- The probe should not self-anneal to form secondary structure and should not be picked from a gene region with high likelihood of secondary structure. Secondary structure formation may reduce hybridization efficiency, leading to reduced assay sensitivity.

- The probe should be selected from gene-specific regions with reasonable sequence complexity to avoid cross-hybridization.
- The probe should be as close to the forward and reverse primers as possible, without overlapping the primer sequences. The amplicon size is usually in the range of 60–150.

### 5.1.12 Molecular beacon probe design

Molecular beacon probes form stem-loop hairpin structures at low temperature. In this closed state, the fluorophore and the quencher are held in close vicinity, and thus no fluorescence signals are detected (Tyagi and Kramer, 1996). The hairpin loop contains gene specific sequence for hybridization in real-time PCR. Thus, the rules for TaqMan® probe design also apply to molecular beacon loop design. The hairpin structure is disrupted by hybridization to the amplicon, which separates the fluorophore from the quencher for fluorescence detection. Because of this unique structural requirement, one major task for molecular beacon probe design is to identify a suitable hairpin structure that melts 7–10°C higher than the PCR primers. The Tm of the hairpin stem cannot be calculated using the Tm formulas for PCR primers because the stem is formed by intramolecular folding. In general, programs for secondary structure prediction, such as Mfold (Zuker, 2003), can be used to predict hairpin stem melting temperature. The stem usually consists of 5–7 base pairs with 75–100% GC content. A G residue should not be placed at the end of the stem because it will quench the fluorophore.

## 5.2 PrimerBank – an online real-time PCR primer database

### 5.2.1 Primer design algorithm

We have developed a real-time PCR primer design algorithm based on the general guidelines described in Section 1 (Wang and Seed, 2003a). An outline of the algorithm is presented in *Figure 5.1*. The algorithm is implemented by a design tool called uPrimer, with which we have designed more than 300,000 primers encompassing most human and mouse genes. The following comprise detailed descriptions of the properties of these primers.

#### Primer specificity

Although primer specificity is one of the most important requirements in real-time PCR, most primer design programs take only one target sequence without considering mispriming to off-target templates. At the present time, to address the mispriming problem, one can design a number of primer pairs and then individually check cross-matches of each primer with BLAST (Altschul *et al.*, 1990). However this screening step is incomplete and does not consider some important design criteria. For example, cross-matches at the 3' end of a primer are more likely to produce non-specific amplicons. In our experience, only about two thirds of the primers designed in the conventional way can be used in real-time PCR experiments.